

1/23

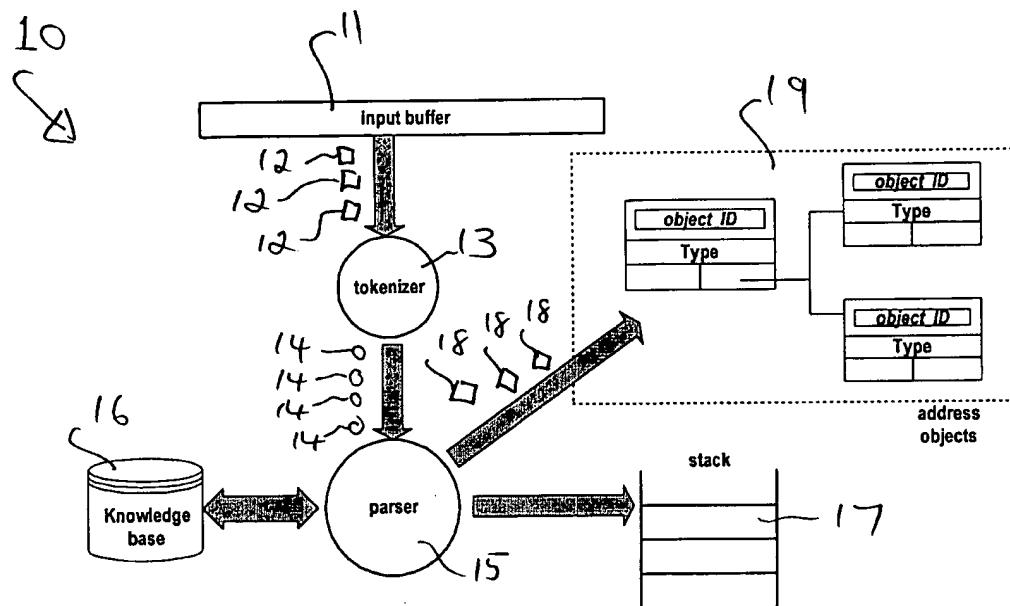


Figure 1. The conceptual parsing architecture.

Input buffer: the data structure that contains the character string to be parsed.
We assume the characters are encoded by UNICODE.

FIG. 1

TO 05 T0 00 - 06 15 01

2/23

10 09 08 07 06 05 04 03 02 01

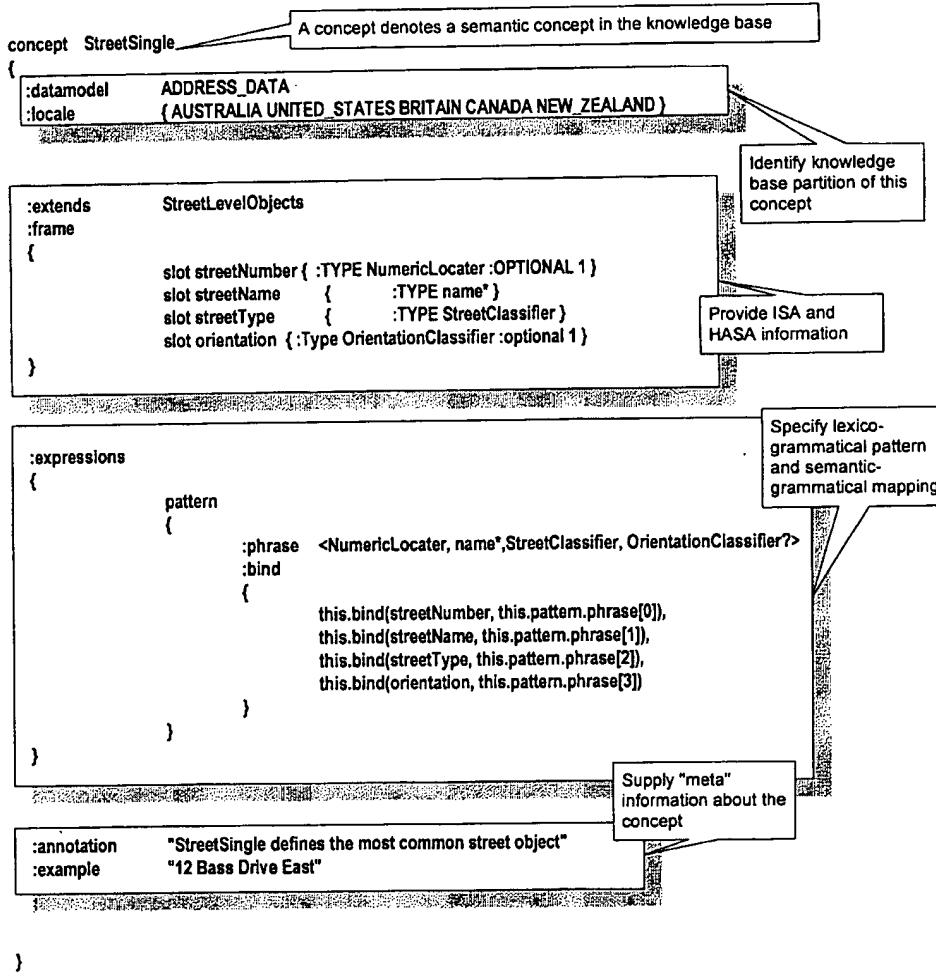


FIG. 2

3/22

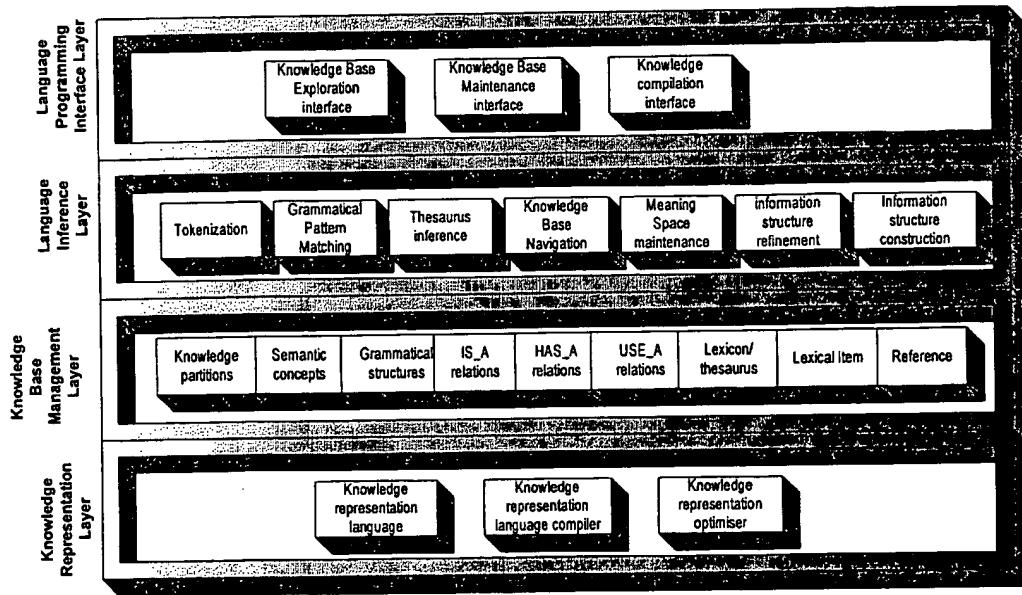


FIG. 3

4/23

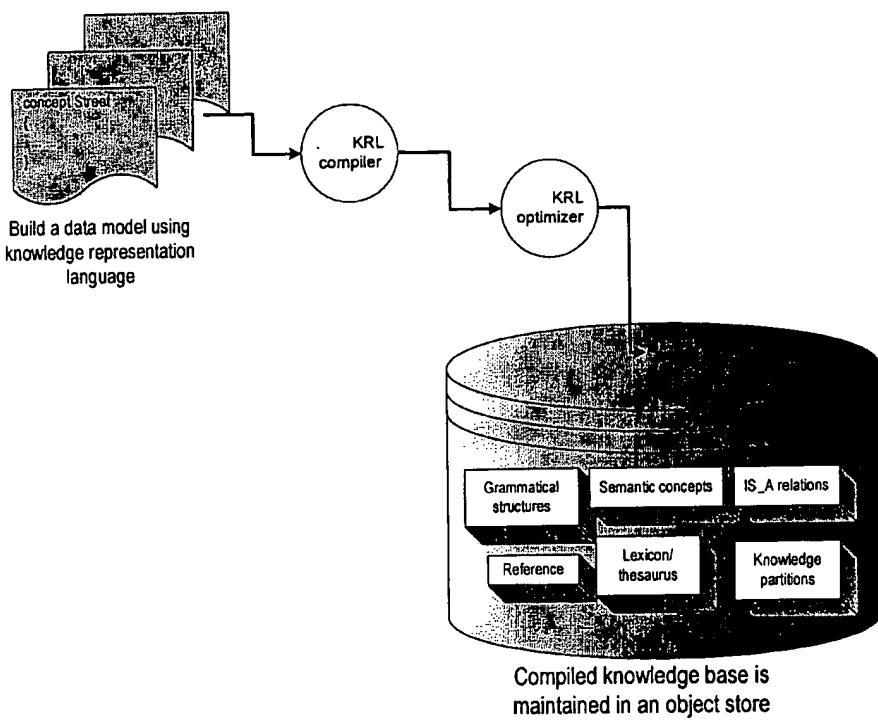


FIG. 4

5/23

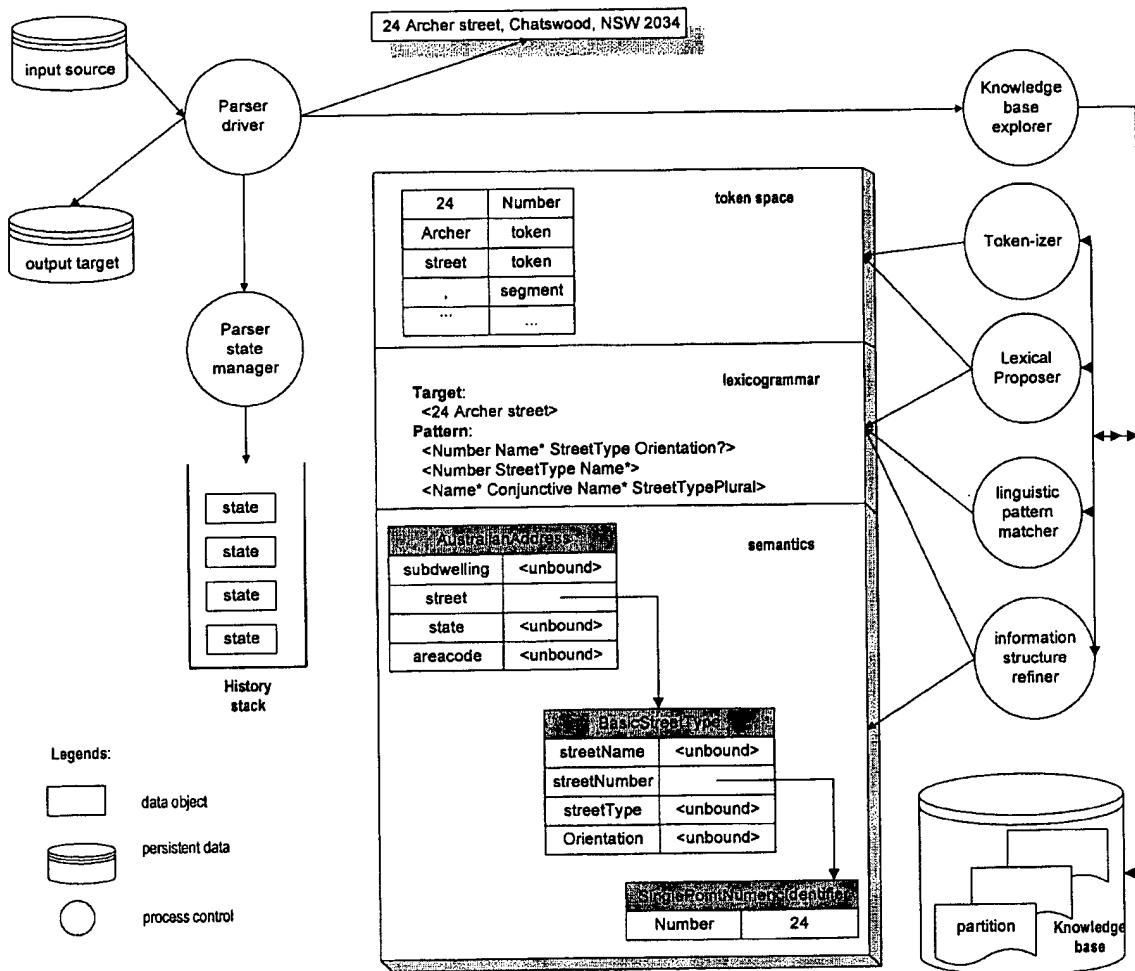


FIG. 5

TO 5790-202-000000

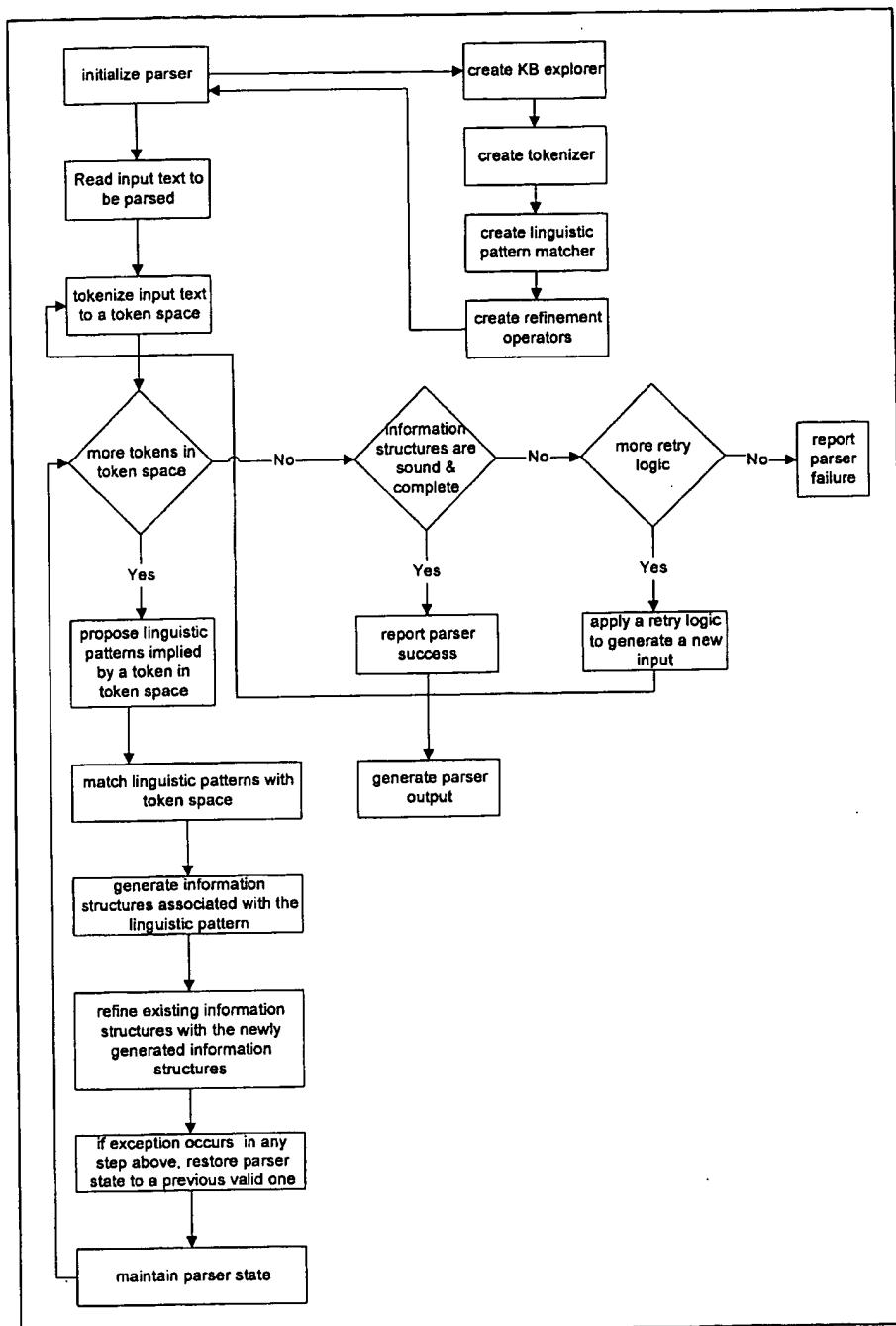


FIG. 6

7/23

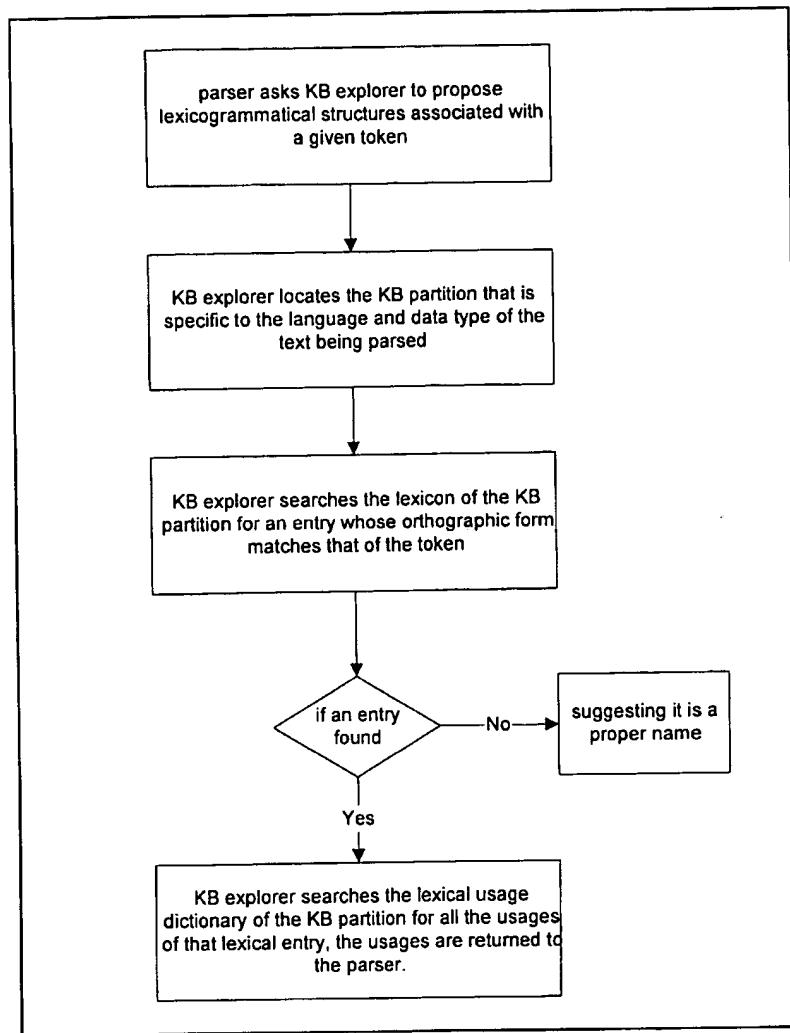


FIG. 7

0908242009 - 064504

8/23

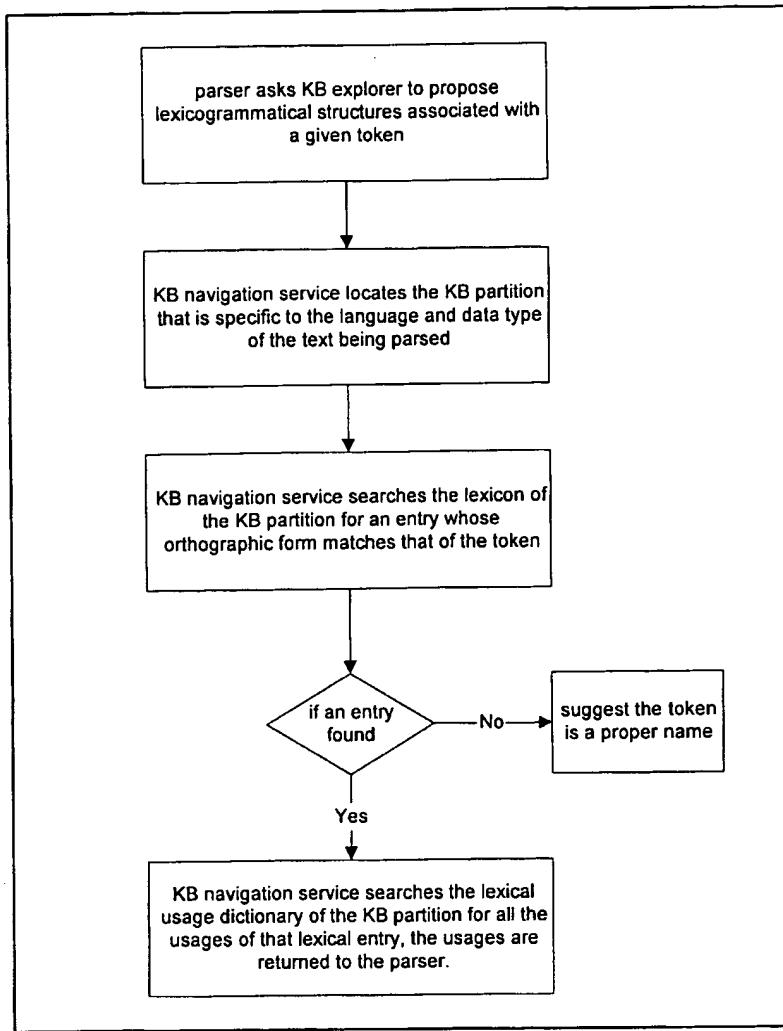


FIG. 8

090823232323-061501

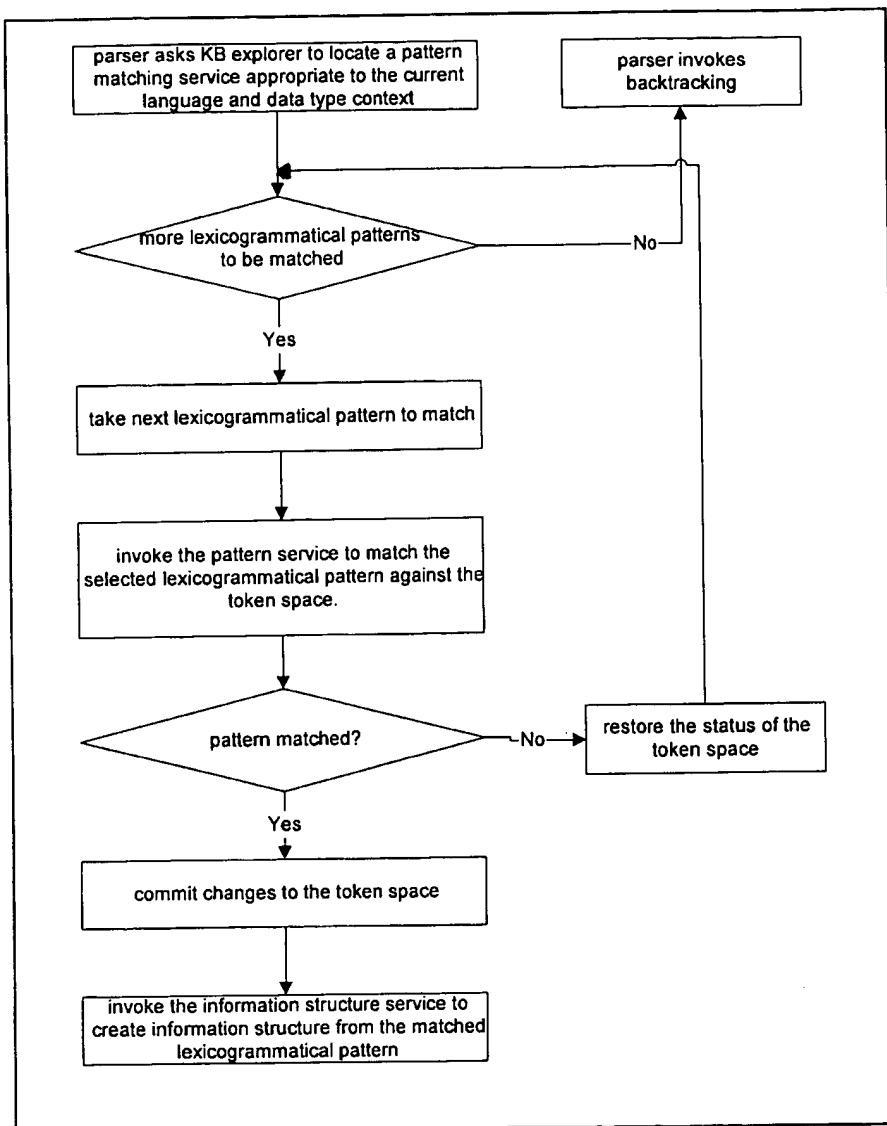


FIG. 9

TOP SECRET//COMINT

10/23

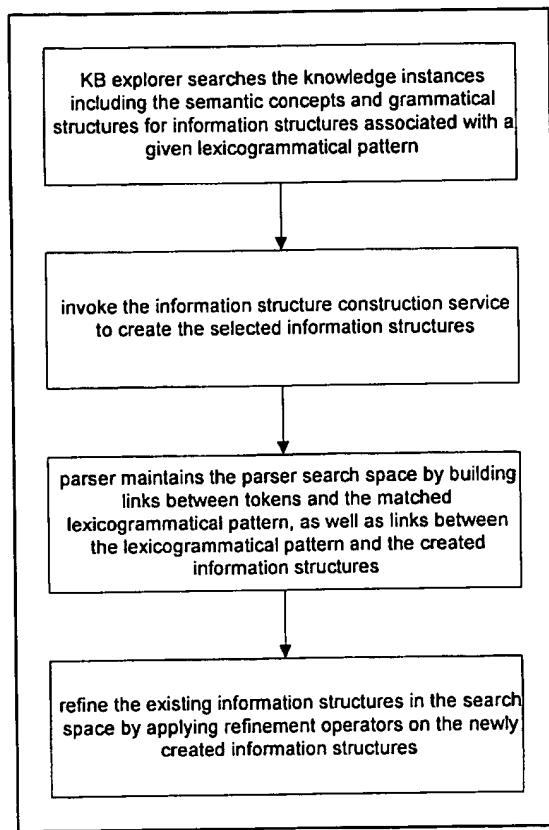


FIG. 10

0988234123 - 061504

11/23

Example address: unit 6, 22 Fontenoy road,

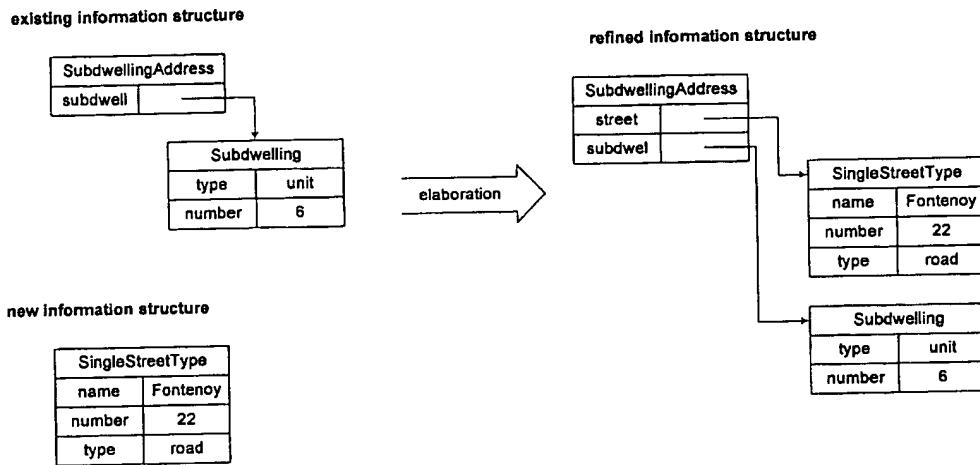


FIG. 11

105700-200700-001

12/23

Example address: Fontenoy road and Curzon street,

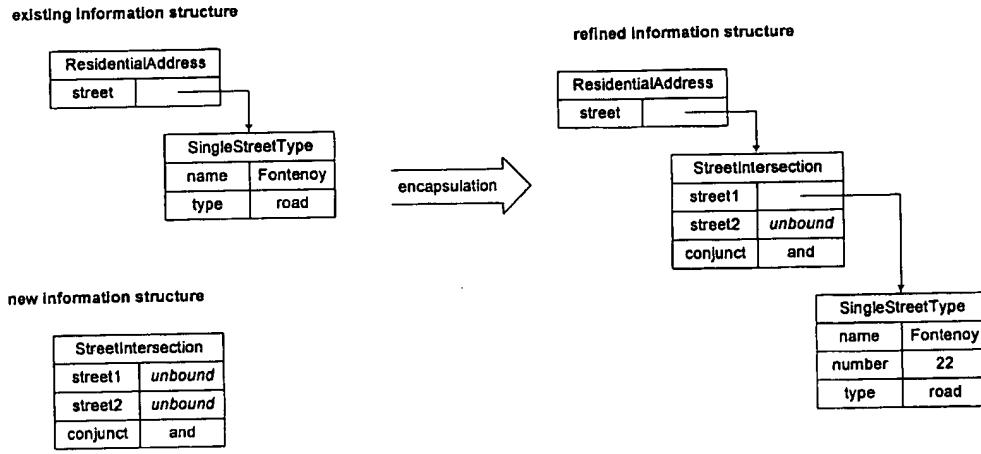


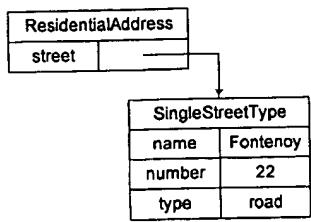
FIG. 12

098342201
TOSTERD

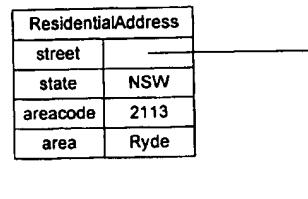
13/23

Example address: 22 Fontenoy road, Ryde, NSW 2113

existing information structure



refined information structure



enhancement →

new information structure

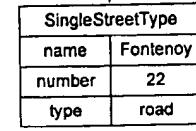
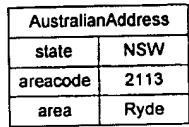


FIG. 13

01234567890123456789012345678901

14/23

Example address: *unit 6, ...*

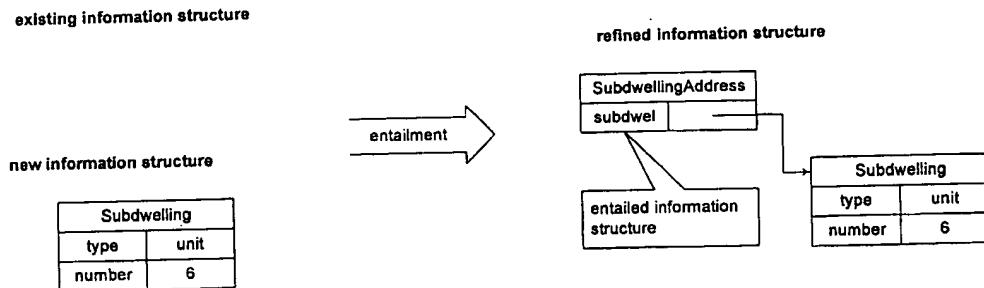


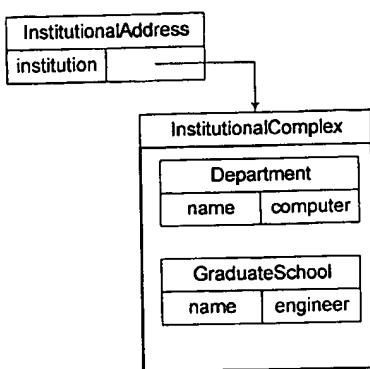
FIG. 14

098834223-061504

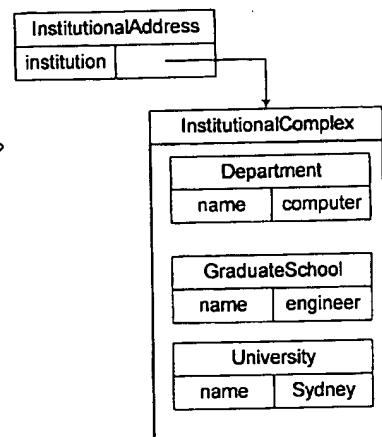
15/23

Example address: Dept. of computer science, school of engineering, Univ. of Sydney,

existing information structure



refined information structure



new information structure

University	
name	Sydney

FIG. 15

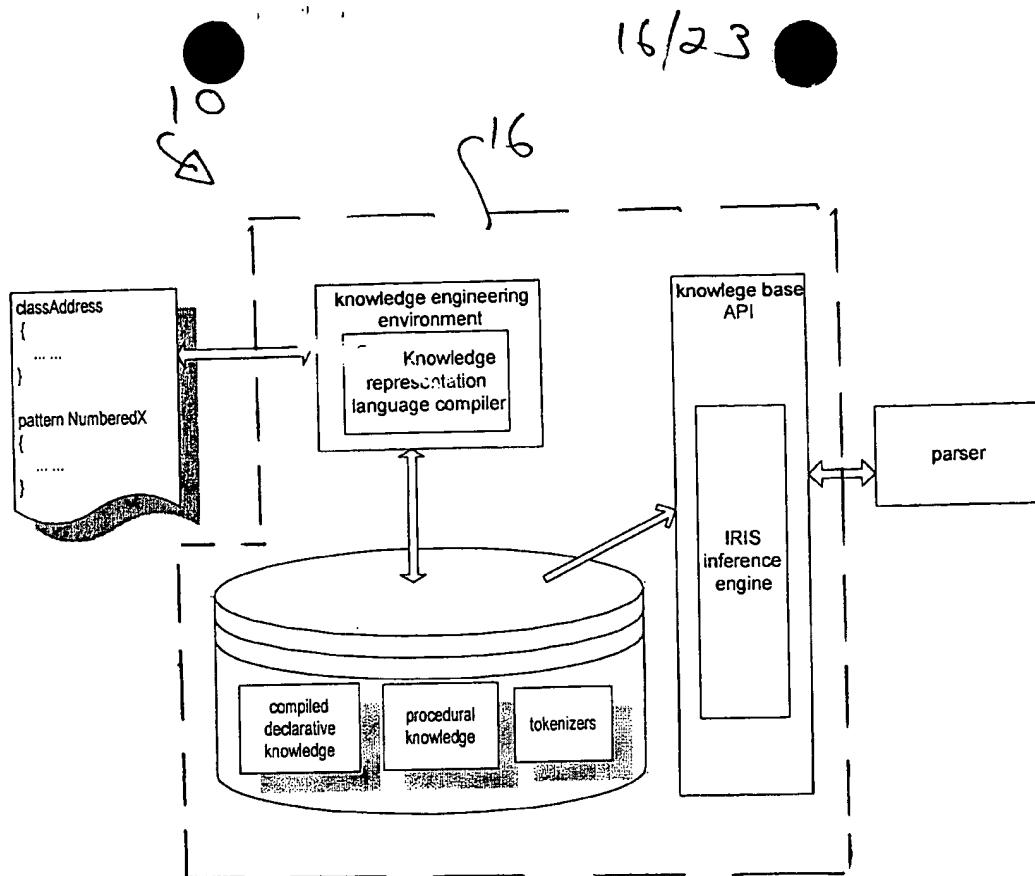


FIG. 16

098831223 061501

17/23

TESTED-CODED-REVIEWED

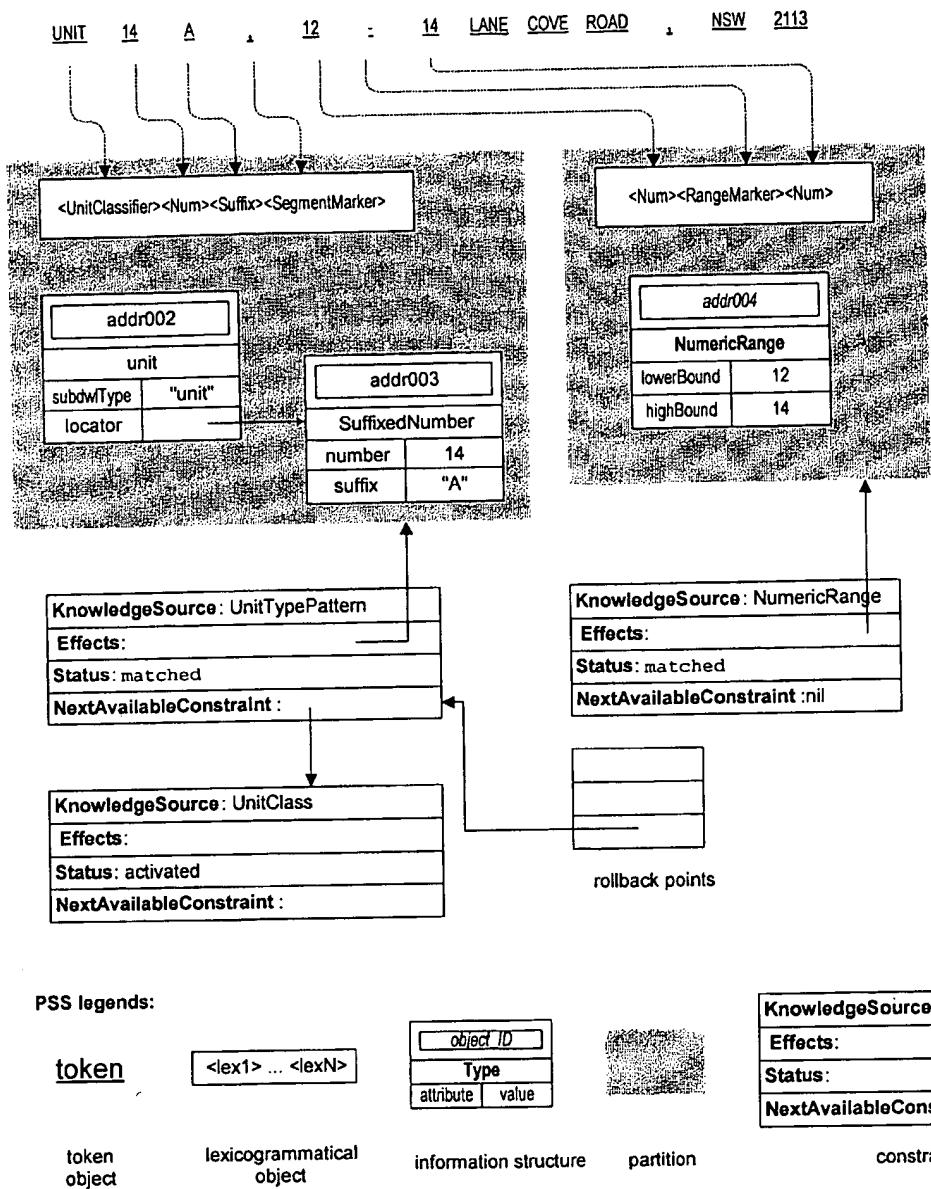


FIG. 17

18/23

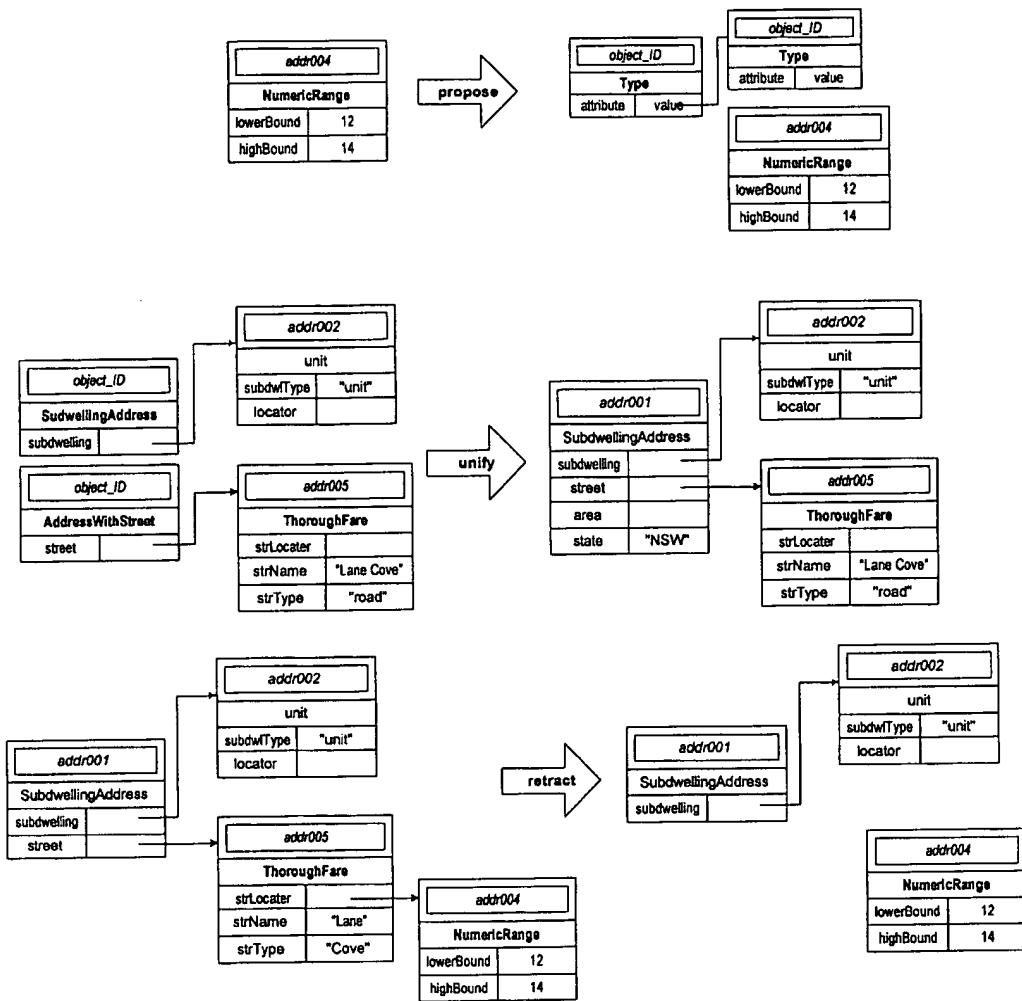


FIG. 18

01054690 02032144 00000000

19/23

Token	lexical type
unit	(subdwellingClassifier)
14	(numeric)
a	(affix)
,	(segment-marker)
12	(numeric)
,	(range-marker)
14	(numeric)
Lane	(thoroughfareClassifier) (properName)
Cove	(thoroughfareClassifier) (properName)
road	(thoroughfareClassifier) (properName)
,	(segment-marker)
Ryde	(properName)
,	(segment-marker)
NSW	(stateName)
2113	(AreaCode)

Figure 19.1 Initial state of parsing.

FIG. 19.1

09833423 064501

2c/23

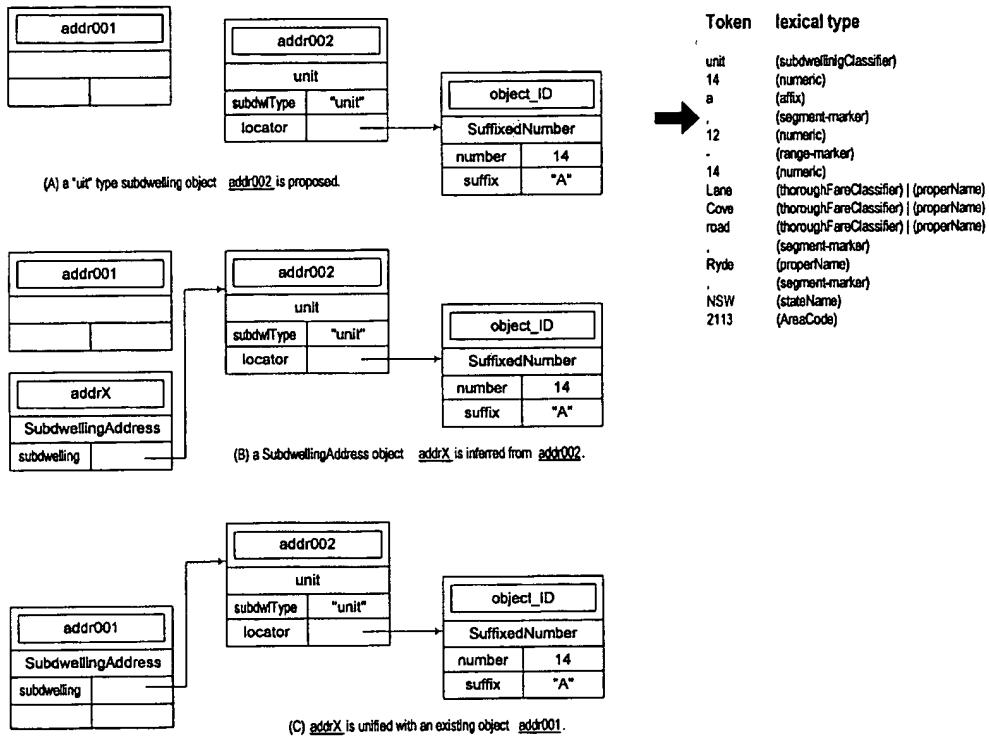


Figure 19.2. Address objects built after parsing "unit 14A".

FIG. 19.2

21/23

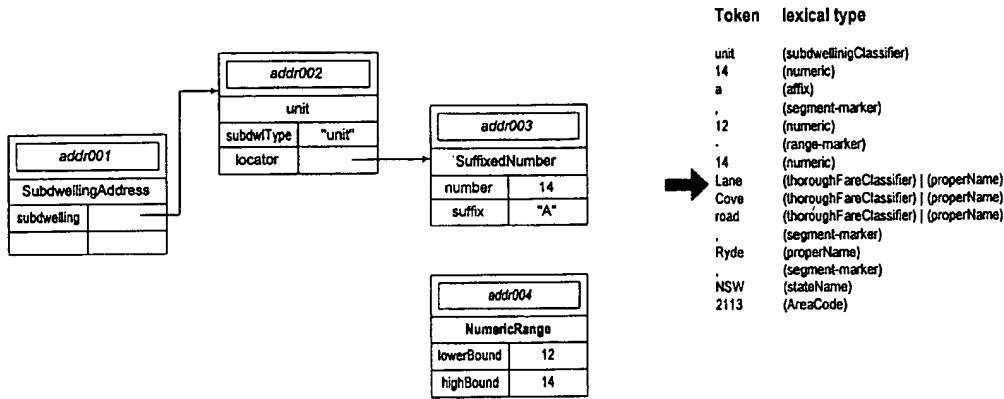


Figure 19.3. Temporary information structure held in stack.

FIG. 19.3

22/23

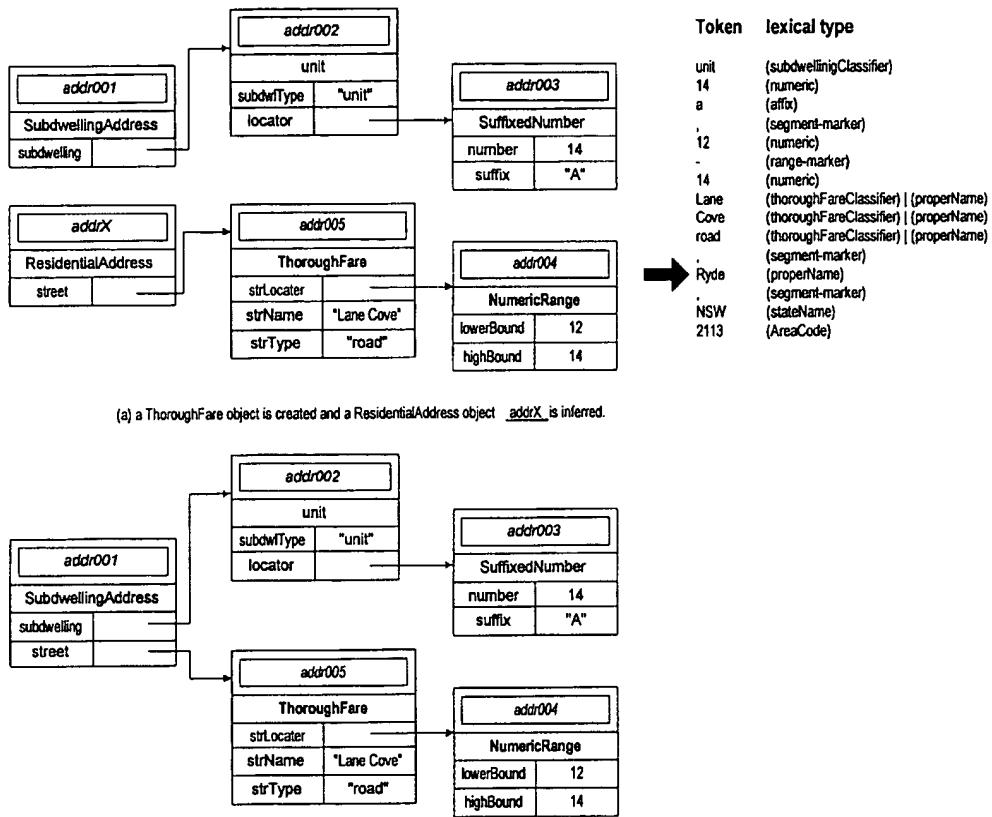


Figure 19.4. Information structure obtained after parsing "12-14 Lane Cove Road".

FIG. 19.4

23/23

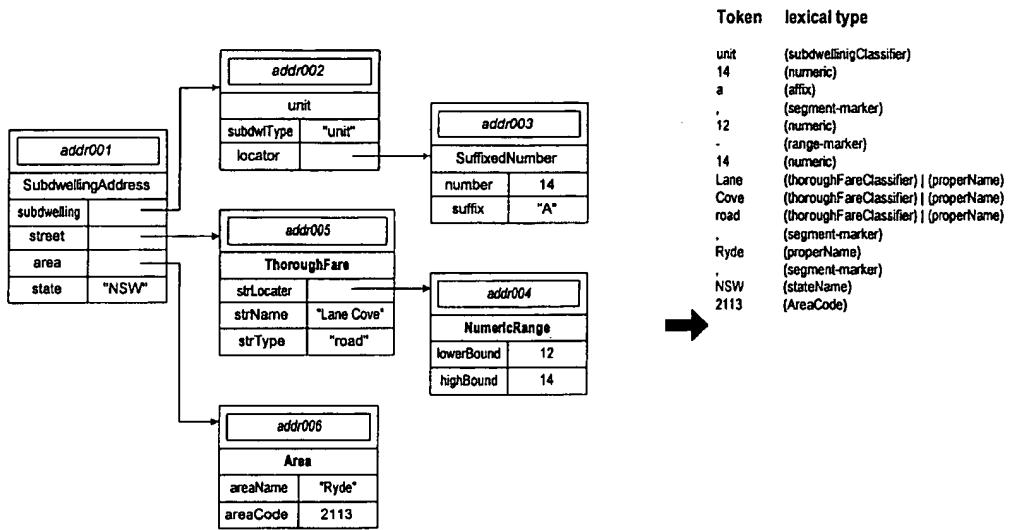


Figure 19.5. The final address information structure.

FIG. 19.5

TOP SECRET - 0641604